Supplementary Materials:

Supplemental Methods

Supplemental Figure Legends S1-S5

Supplemental Tables Descriptions S1-S14

Supplemental Methods Figure Legends SM1-SM15

Supplemental References 1-7

Supplemental Methods Figures SM1-SM15

Supplemental Methods:

Whole exome capture and sequencing. DNA was processed for sequencing as previously described (1). Briefly, high molecular weight DNA was sheared to 300 bp and pre-capture libraries were generated and amplified over 6-8 cycles by PCR. DNA was denatured and annealed in liquid phase to SeqCap EZ Exome or VCRome 2.1 target probes (Roche NimbleGen, Madison, WI). The bound fragments were amplified with a 12-cycle post-capture linker-mediated PCR. The efficiency of the capture was evaluated by performing a qPCR-based quality check on four standardized oligo sets (RUNX2, PRKG1, SMG1, and NLK) The enrichment of the controls in the tumor and blood capture libraries was estimated to range from 7 to 10 fold over the background based on qPCR. The 2 captured libraries were further processed for Illumina sequencing. Sequencing reactions were extended for 202 cycles of single base sequencing on an Illumina HiSeq 2000 sequencing instrument according to the manufacturer's instructions. Real Time Analysis (RTA) software was used image processing and base calling. On average, about 80-100 million successful reads, consisting of 2x100 bp, were generated on each lane of a flow cell. All reads that passed the illumina quality filter were formatted into fastq

files. The fastq files are aligned to the genome using BWA (bwa-0.5.9rcl) against human reference genome build36. Default parameters are used for alignment except for a 40 bp seed sequence, 2 mismatches in the seed, and a total of 3 mismatches allowed.

Alignment and Variant Calling. The result of each exome alignment was stored in BAM (Binary Alignment Map) format. BAM files generated from alignment of Illumina sequencing reads were preprocessed using GATK (2) to recalibrate base qualities and refine alignments neighboring insertions or deletions (indels). Single nucleotide variants and indels were identified as previously described (3). Mutations are validated by sequencing PCR amplified DNA from the mutated sample and its matched normal on a 454 instrument. The mutation rate for each gene was computed across the population and corrected for its length and base composition using the Mutational Significance in Cancer package (MuSiC: http://gmt.genome.wustl.edu/genome-music/current/). We accepted all genes as significantly mutated whose P-value, corrected for multiple tested yielded a False Discovery Rate of less than 0.1.

mRNA expression arrays. Total RNA was isolated from tumor tissue by using Tri-reagent and hybridized to Affymetrix GeneChip Human Exon 1.0ST Arrays (Affymetrix) according to manufacturer's instructions. Cel files were analyzed in R (v2.12.0) with affy, gdata, aroma.affymetrix, NCStats, ClassDiscovery, gplots, and annotationTools packages. Data was background corrected with RMA and quantile normalized. NUSE plot, hierarchical clustering, and PCA of normalized data are shown in Fig. SM1-SM3. Samples 391 and 258 were removed due to poor data quality. The resulting exon-level data were converted to transcripts using Core

probesets definition (HuEx-1_0-st-v2,coreR3,A20071112,EP.CDF which contains 18,708 units/transcript clusters, 284,258 groups/probesets, and 1,082,385 probes) for further analysis.

MicroRNA expression arrays. Total RNA was isolated from tumor tissue by using Tri-reagent and hybridized to Agilent Human miRNA microarray rel12.0 (Agilent) according to manufacturer's instructions. Data were analyzed in R with AgiMicroRna, limma, gplots, xtable, ClassComparison, ClassDiscovery, Hmisc and affy packages. Control spots and probes without detectable signal were removed, 432 probes remained. Data was background corrected with RMA and quantile normalized. An estimate of the microRNA gene signal was obtained by fitting a linear model to log2 transformed probe intensities. This model produces an estimate of the gene signal taking into account the probe effect. The model parameters estimates are obtained by the median polish algorithm. Batch effects were removed by adjusting means. Hierarchical clustering and PCA of the adjusted data are shown in Fig. SM4-SM5.

SNP arrays. DNA was hybridized to Affymetrix SNP6.0 arrays (Affymetrix) according to manufacturer's instructions. Segmentation was performed by using Partek software (v6.6, Partek Inc.). CEL and CHP files were loaded into Partek to generate the signal intensity file. Data were adjusted for GC content. Allele specific summarization was performed. Copy number estimates were created from intensities. Samples were paired based on the subject IDs, and normal samples were used as baseline. Segmentation was performed on the copy number file with the following parameters:

. minimum genomic markers: 20

. two sided t-test p-value threshold: 0.0001

. signal to noise ratio: 0.4

. region report cutoff: segment means fall out of 1.7 and 2.3

Segments were associated with genes in that genomic location. Fig. SM6 shows a PCA of the SNP data.

Segmentation data were exported from Partek and loaded into R for ASCAT analysis (v1.0). Tumor percentage (TP) and ploidy were estimated by using ASCAT. Copy numbers in the segmented data were adjusted by the TP value from ASCAT by using the following formula: (orig.seg.mean - 2)/TP + 2. For example, sample A has 50% tumor. One of its segments has copy number 1.5. After adjustment, the new segment mean is 1 ((1.5-2)/0.5+2=1). Next the copy number was adjusted by the ploidy value from ASCAT with the following formula: TP.adj.mean + ploidy -sum(TP.adj.mean * length.of.that.segment)/sum(all.segment). Fig. SM7 shows the distribution of copy numbers in all samples before and after the ASCAT adjustment. The TP and ploidy values were further adjusted so that the majority of the copy numbers would be close to an integer value using the following method:

- identify major modes in the adjusted copy number data based on data density.
- Minimize the Mean Square Error (MSE) between major modes and the closest integers.

Negative adjusted mean values were truncated at 0. The Copy Neutral Event (CNE) was defined as the closest integer that has the highest density of segments. Amplifications and deletions were then defined as CNE+0.5 and CNE-0.5, respectively. Fig. SM7-SM8 show the distribution of copy numbers in all samples before and after the final adjustment. Samples 166, 258, 350, 562, 530 were removed due to poor data quality or mismatched tumor and normal specimens.

Relative CN was defined relative to the CNE, while absolute CN was defined as the actual number of calculated copies in each region.

Methylation arrays. DNA was prepared from 40 tumors, 2 normal oral epithelial tissues, and 2 normal blood samples and hybridized to Illumina HumanMethylation450 arrays according to manufacturer's instructions. Methylation beta-values were exported from the Illumina software. The Illumina annotation file was used for methylation probe site information. Probes on X and Y chromosomes were removed. Probes with a beta-value >=0.25 in at least 2 normal samples were removed. Probes were also required to have a beta-value >=0.25 in at least 2 tumor samples and annotated as CpG island for further analysis.

To detect a bimodal methylation pattern, we first transformed the methylation beta-value across all samples, based on $y = \log2(x/(1-x))$, where x is the beta value. We then applied the Bimodality Index algorithm, devolved in house, and generated a bimodal index value for each methylation probe (4). Each sample was assigned to a methylation group (methylated or unmethylated) by using the model-based clustering algorithm, Mclust as an R package in Bioconductor, developed by Fraley and Raftery (5). We assume equal variance in both groups. Detailed description about the algorithm can be found in the reference. A Bimodality Index cutoff of 2.3 was used to generate the probe list for the clustering. Unsupervised hierarchical clustering was performed on methylation group data by using Spearman correlation and Ward's linkage. Two visualizations of the clustering were generated, one showing the sample classification matrix (class1 and class 2; or methylated and unmethylated (Fig 2C); the other showing the centered transformed beta-values (Figure SM9).

SNP mRNA integration. Thirty-seven samples generated data on both SNP and mRNA expression platforms and these data were integrated. Each gene in each sample was assigned both a relative and absolute CN level from the SNP data described above. Absolute CN is the calculated number of copies for that gene. Relative CN is the difference between the absolute CN for that gene and the CNE of the sample. Relative and absolute CN values were rounded to the nearest integer copy number. If a gene was found to have different CNs for different parts of the gene then the lowest of the values was used as the CN level. Linear regression was used to test the relationship between CNV levels and mRNA expression values. Only genes that occur in at least 2 CN levels were analyzed. BUM plots for the correlations between gene expression and relative or absolute CN are shown in Fig. SM10-SM11, respectively. An example of the difference between relative and absolute CN correlations with gene expression is shown in Fig. SM12-SM13.

Copy and expression variants (CEV): We will refer to each gene in each sample as an event. All events were categorized into 1 of 3 CNV levels (loss, unchanged, or gain) based on the relative CN. The combination of 15783 genes and 37 samples resulted in 583971 events for analysis. “Loss” was any event with a relative CN <-0.5, “gain” were events with a relative CN >0.5, and “unchanged” were events with a relative CN between -0.5 and 0.5. For each gene, 95% confidence intervals (CI) of gene expression were calculated based on all samples an unchanged CN for that gene. These CI were then used to put the gene expression for each event into 1 of 3 levels (down, normal, or up). “Down” was expression below the CI, “up” was expression above the CI, “normal” was expression within the CI. These calls for CN and gene expression

generated a matrix of 9 possible categories for each event. Fig. SM14 shows the distribution of all events into the 9 categories. CEV were defined as events with “loss” and “down” or “gain” and “up”.

P16 expression was calculated by using the average of probesets 3201466 and 3201467. Only 5 samples did not demonstrate a CN loss or DNA methylation at CDKN2A. These samples were not sufficient to calculate CEV for p16, therefore for pathway analysis involving p16 all copy number losses were considered genomic alterations at this locus.

SNP microRNA integration. Thirty-seven samples generated data on both SNP and mRNA expression platforms and these data were integrated as described above for mRNA. BUM plots for the correlations between microRNA and relative or absolute CN are shown in Fig. SM15.

mRNA and microRNA clustering. To perform unsupervised hierarchical clustering of RNA expression the 2500 genes with the most variable expression were chosen based on the standard deviation. All 432 variable microRNAs were also included. These 2932 genes and microRNAs were clustered by using Pearson correlation and Ward’s linkage.

IHC. Sections (5 microns) cut from formalin fixed archival specimens were deparaffinized and hydrated by successive incubations in xylene, 100% ethanol, and 95% ethanol followed by rinsing in water. Following antigen retrieval in Citrate (pH 6), immunohistochemistry (IHC) was performed with a Lab Vison autostainer 360. Essentially slides were blocked, incubated with primary antibodies (60 min at room temperature) and signal detected with an UltraVision LP

detection system utilizing an HRP Polymer and Diaminobenzidine substrate, followed by counterstaining with hematoxylin. The anti-RPS6KB1 (SC-230) antibody used at 1:150 and anti-Jagged 1 (SC-8303) antibody used at 1:50 were both from Santa Cruz Biotechnology.

TCGA data. TCGA copy number analysis was collected from the Broad TCGA Firehose output. (HNSCC-gdac.broadinstitute.org_HNSC.CopyNumber_Gistic2.Level_4.2012062300.0.0) (Colorectal-gdac.broadinstitute.org_COADREAD.CopyNumber_Gistic2.Level_4.2012062300.0.0) Total CNVs were calculated as the sum of all focal copy number gains and losses in each sample in the regions identified by GISTIC. HNSCC mutation calls were provided as part of the HNSCC TCGA Analysis Working Group.

Targetable genes. A list of druggable cancer-related genes was compiled from an NCI website of targeted therapies (http://www.cancer.gov/cancertopics/factsheet/Therapy/targeted) and an online therapeutic target database (http://bidd.nus.edu.sg/group/cjttd/) assembled by the Bioinformatics and Drug Design group (6), National University of Singapore, Singapore. Only drugs with FDA approval or in human clinical cancer trials within the last 2 years were included. Targets for individual drugs were confirmed by consulting the NCI drug dictionary website (http://www.cancer.gov/drugdictionary). Many small molecule kinase inhibitors in use are promiscuous with respect to target spectrum because they interact with the ATP binding site of enzymes. Therefore, additional targets for selected drugs were identified from a published comprehensive in vitro kinase screen (7), with the criterion that drugs inhibit these additional kinases with an efficacy similar to or better than the canonical targets listed on the NCI drug

dictionary website. The final list of druggable cancer-related genes was generated to include only targets for which our genomic analysis detected either a mutation or CEV in at least one patient tumor.

Supplemental Figures:

Fig. S1. Relative copy number changes in OSCC. A) CIRCOS plot of relative copy number changes in 38 OSCC tumors. Chromosome bands and locations are shown in the outside ring. The next ring is a plot showing the sum of copy number gains and losses for all samples. Red indicates gains and blue indicates losses, and the size indicates how many samples have that change. The rest of the rings indicate the copy number alterations for each sample, and the intensity of color indicates the level of copy number change. B) GISTIC analysis of focal copy number alterations. Copy number gains are shown on the left (red) and losses on the right (blue). The size of each peak indicates the FDR Q-value with the top axis for gains and the bottom axis for losses. Significant regions (Q-value <0.25) are indicated by the cytoband location.

Fig. S2. Integrated analysis of gene expression and copy number alterations. A) Chromosome arm-level copy number alterations. Q-values were generated by GISTIC. Frequency is based on the number of samples with the indicated copy number change to >50% of the chromosome arm. B) Comparison of expression and copy number (CN) correlation p-values between absolute CN and relative CN. 224 genes with significant p-values on both comparisons are shown. The horizontal line indicates where a perfect association would be found. The majority of the genes are below the line indicating a lower p-value for the relative CN association. C) Immunohistochemical staining (IHC) demonstrates high protein expression in samples with a copy number variant and high expression. IHC was performed for the indicated proteins. Sample

number is indicated in the upper right corner and relative CN is indicated in the lower left corner. All samples with a relative CN >0 had high expression for the indicated gene.

Fig. S3: Full length NOTCH1 expression in the cell line HN4. A) Western blot for cleaved or total NOTCH1 protein. FL indicates full length NOTCH1, and TM indicates the transmembrane form. Actin is shown as a loading control. HN4-NFL indicates cells expressing full-length NOTCH1 and HN4-MigR1 indicates the vector control cells. Cells were plated on coated dishes to determine if ligation of NOTCH1 with ligand would alter growth. Nc indicates non-coated plates, Fc indicates plates only coated with antibody Fc and Jag1 indicates plates coated with recombinant ligand Jagged1. Quantitation of bands is shown on the right. B) Relative cell fractions for the "horse-race" experiment. GFP-positive fraction was normalized to the frequency at day 4. Vector controls are shown with dashed lines and NFL-expressing cells are shown with dashed lines. Shapes and colors indicate growth on different substrates as above in A.

Fig. S4: Induction of cell cycle arrest and senescence by activated NOTCH1. A) Phase contrast and fluorescent images of HN31 cells taken 5 days post infection with virus to express ICN1 or empty vector. Cells were sorted for GFP positivity and plated and similar densities on day 3 post infection. B) Quantitation of senescence associated β-galactosidase staining from Fig. 3H. Greater than 400 cells were counted for each cell type in duplicate. P-value was calculated with t-test on transformed values.

Fig. S5: Inhibition of CASP8 promotes tumorigenicity. A) HN4 cells were infected with the indicated construct and selected with puromycin. Western blot for caspase-8 protein indicates dramatically reduced caspase-8 levels with the shCASP8-30 construct. B) The indicated cells were injected into mouse tongues and tumor size was measured over time. shCASP8-30 resulted in larger tumors when compared to the other constructs.

Supplemental Tables:

Table S1. Clinical information

Table S2. Clinical and genomic variables

Table S3. GISTIC peaks

Table S4. GISTIC events in each sample

Table S5. mRNA expression and copy number correlations

Table S6. microRNA expression and copy number correlations

Table S7. Differential expression between expression clusters 1 and 2

Table S8. Methylation heatmap values

Table S9. Sequencing coverages

Table S10. Mutation significance

Table S11. Mutation data (MAF)

Table S12. Cell line NOTCH1 and CASP8 mutations

Table S13. Pathway events

Table S14. Targetable gene events

Supplemental Methods Figures

Fig. SM1: Normalized Unscaled Standard Error (NUSE) plot of the log2 transformed transcript-level normalized data. Plot was generated in R from the expression array data. Larger variation indicates poorer sample quality.

Fig. SM2: Hierarchical clustering using correlation distance metric and wards linkage of the log2 transformed transcript-level normalized data. Samples were not found to cluster based on any of the shown variables, indicating a lack of batch effects.

Fig. SM3: Principle component analysis of the log2 transformed transcript-level normalized data. Sample 391 is a possible outlier sample in this and other QC analysis and was removed from the final analysis.

Fig. SM4: Hierarchical clustering using Pearson correlation distance metric and Ward's linkage on log2 normalized microRNA expression data after removing the batch effect. No outliers were identified.

Fig. SM5: Principle component analysis of log2 transformed normalized microRNA expression data after removing the batch effect. No outliers were identified.

Fig. SM6: Principle component analysis of paired tumor-normal SNP data. Sample 166 was identified as an outlier in principle component one and removed from the final analysis.

Fig. SM7: Distribution of copy numbers for all segments in all samples before and after ASCAT tumor percentage adjustment. Blue lines indicate integer values.

Fig. SM8: Distribution of final adjusted copy numbers for all segments in all samples. Blue lines indicate integer values.

Fig. SM9: Clustered heatmap of bimodal methylation analysis. Samples are clustered as in Figure 3C, but transformed beta values are shown instead of the methylation status call.

Fig. SM10: BUM plot for the relationship between relative CN and gene expression. Superimposed curves represent the fit of a beta-uniform-mixture model. Plot indicates enrichment for genes with low p-values, consistent with a significant finding.

Fig. SM11: BUM plot for the relationship between absolute CN and gene expression. Superimposed curves represent the fit of a beta-uniform-mixture model. Plot indicates enrichment for genes with low p-values, consistent with a significant finding, however this plot has a broader curve than the one in Fig. SM10.

Fig. SM12: Expression and relative CNV plot for an example gene (MELK) across all samples.

Fig. SM13: Expression and absolute CNV plot for an example gene (MELK) across all samples. The p-value is lower than for relative copy number in Fig. SM12.

Fig. SM14: Copy and expression variant analysis frequency of events. Number of events in each category is indicated. Each gene in each sample is an event. Copy number categories are indicate across the top and expression categories relative to the 95% confidence interval for all samples with no copy number change. 13418 events are indicated to have a copy number gain and high expression.

Fig. SM15: Histogram showing the distribution of p-values from the relationship between CN and microRNA expression. Superimposed curves represent the fit of a beta-uniform-mixture model. A very modest enrichment for low p-values is observed, suggesting a statistically weak relationship.

Supplemental References:

1. Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature. 2012.
2. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297-303.
3. Network TCGA. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487:330-7.
4. Wang J, Wen S, Symmans WF, Pusztai L, Coombes KR. The bimodality index: a criterion for discovering and ranking bimodal signatures from cancer gene expression profiling data. Cancer Inform. 2009;7:199-216.
5. Fraley C, Raftery AE. Model-Based Clustering, Discriminant Analysis, and Density Estimation. J Am Stat Assoc. 2002;97:611-31.
6. Zhu F, Shi Z, Qin C, Tao L, Liu X, Xu F, et al. Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. Nucleic Acids Res. 2012;40:D1128-36.
7. Anastassiadis T, Deacon SW, Devarajan K, Ma H, Peterson JR. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. Nat Biotechnol. 2011;29:1039-45.

Fig. SM1.

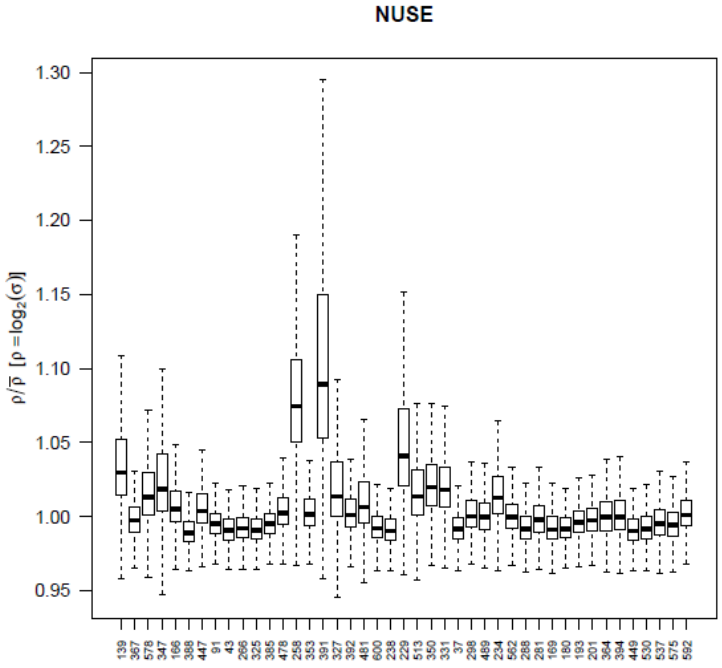

NUSE
1.30
1.25
1.20
1.15
1.10
1.05
1.00
0.95
139 367 578 347 166 388 447 91 43 286 325 365 475 258 353 391 327 392 481 600 238 229 513 350 331 37 298 489 234 562 288 281 169 180 193 201 364 344 449 530 537 575 592

Fig. SM2.

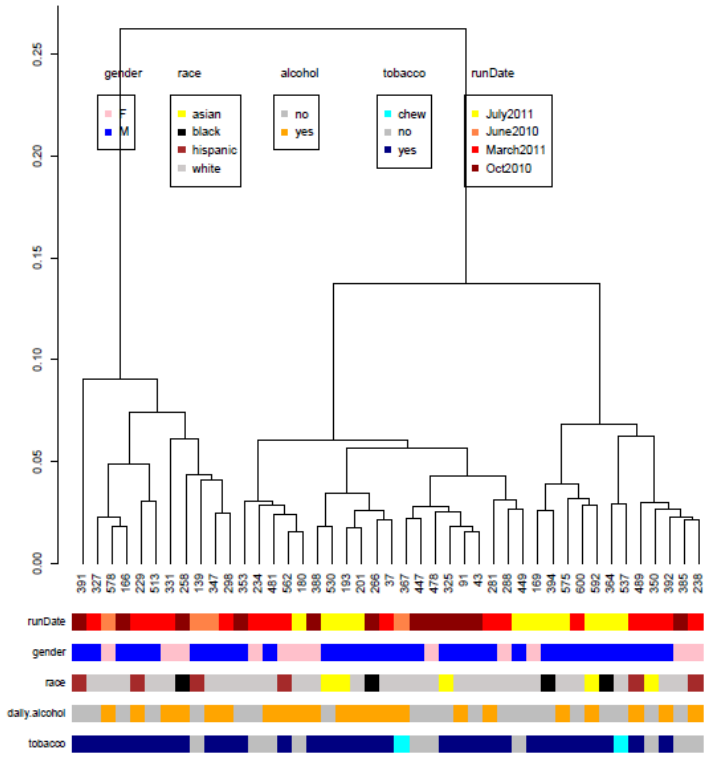

gender
race
alcohol
tobacco
runDate
F
M
asian
black
hispanic
white
no
yes
chew
no
yes
July2011
June2010
March2011
Oct2010
0.25
0.20
0.15
0.10
0.05
0.00
391 327 578 196 229 513 331 258 139 347 298 353 234 481 562 180 388 530 193 201 266 37 367 447 478 325 91 43 281 288 449 169 394 575 600 592 364 537 489 350 392 385 238
runDate
gender
race
daily.alcohol
tobacco

Fig. SM3.

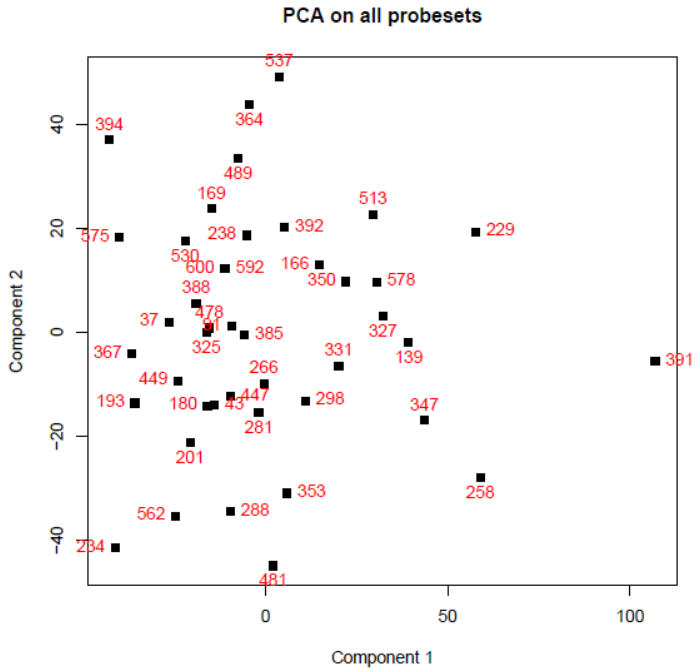

PCA on all probesets
Component 2
Component 1
40
20
0
−20
−40
0
50
100
537
394
364
489
169
513
575
238
392
229
530
600
592
166
350
578
388
478
37
385
327
367
325
331
139
391
266
449
193
180
447
298
347
281
201
353
258
562
288
234
481

Fig. SM4.

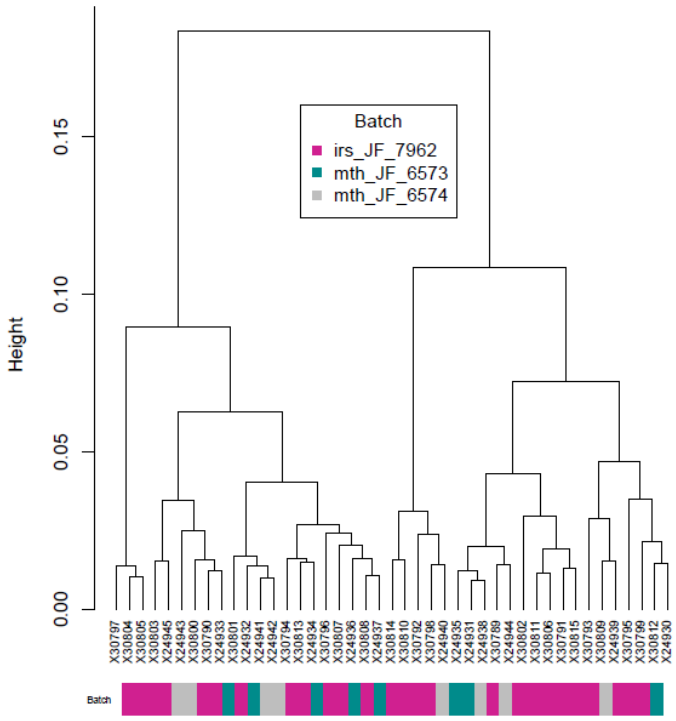

Batch
irs_JF_7962
mth_JF_6573
mth_JF_6574
Height
0.00
0.05
0.10
0.15
Batch

Fig. SM5.

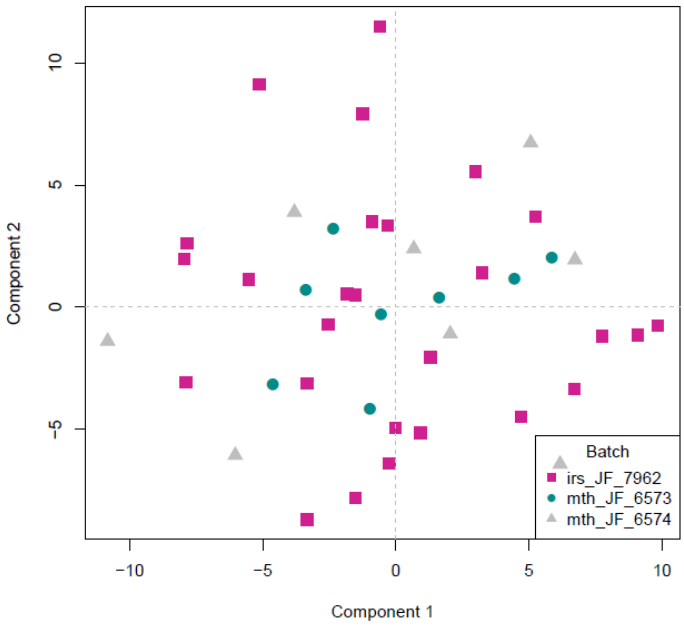

Component 2
10
5
0
−5
−10
−5
0
5
10
Component 1
Batch
irs_JF_7962
mth_JF_6573
mth_JF_6574

Fig. SM6.

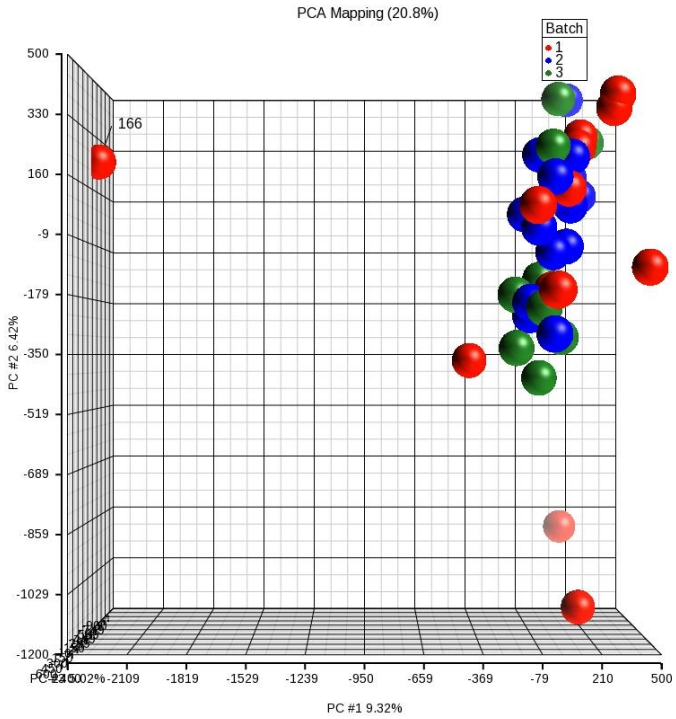

PCA Mapping (20.8%)
Batch
1
2
3
166
500
330
160
-9
-179
-350
-519
-689
-859
-1029
-1200
PC #2 6.42%
-2109
-1819
-1529
-1239
-950
-659
-369
-79
210
500
PC #1 9.32%

Fig. SM7.

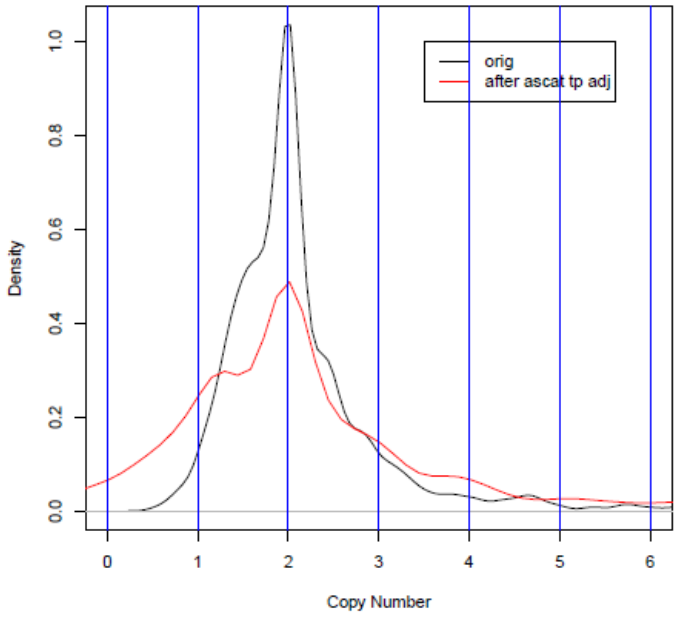

orig
after ascat tp adj
Density
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
Copy Number

Fig. SM8.

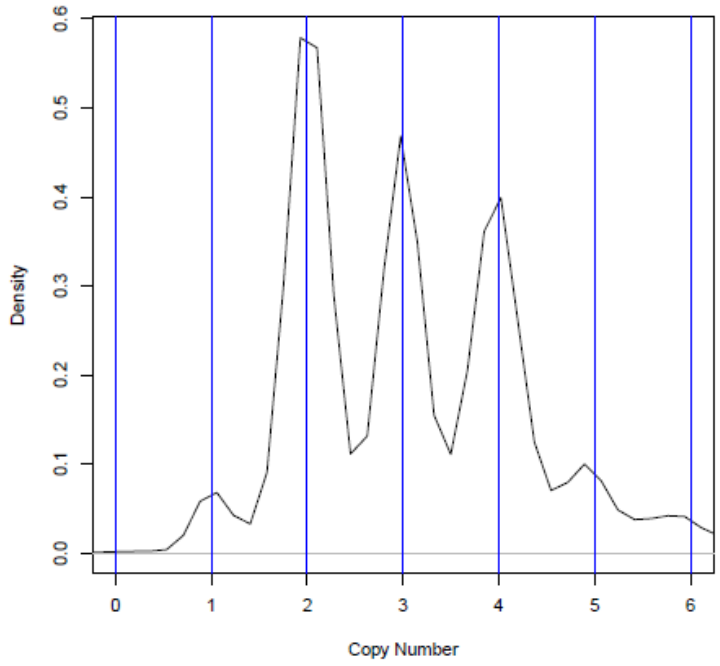


Fig. SM9.

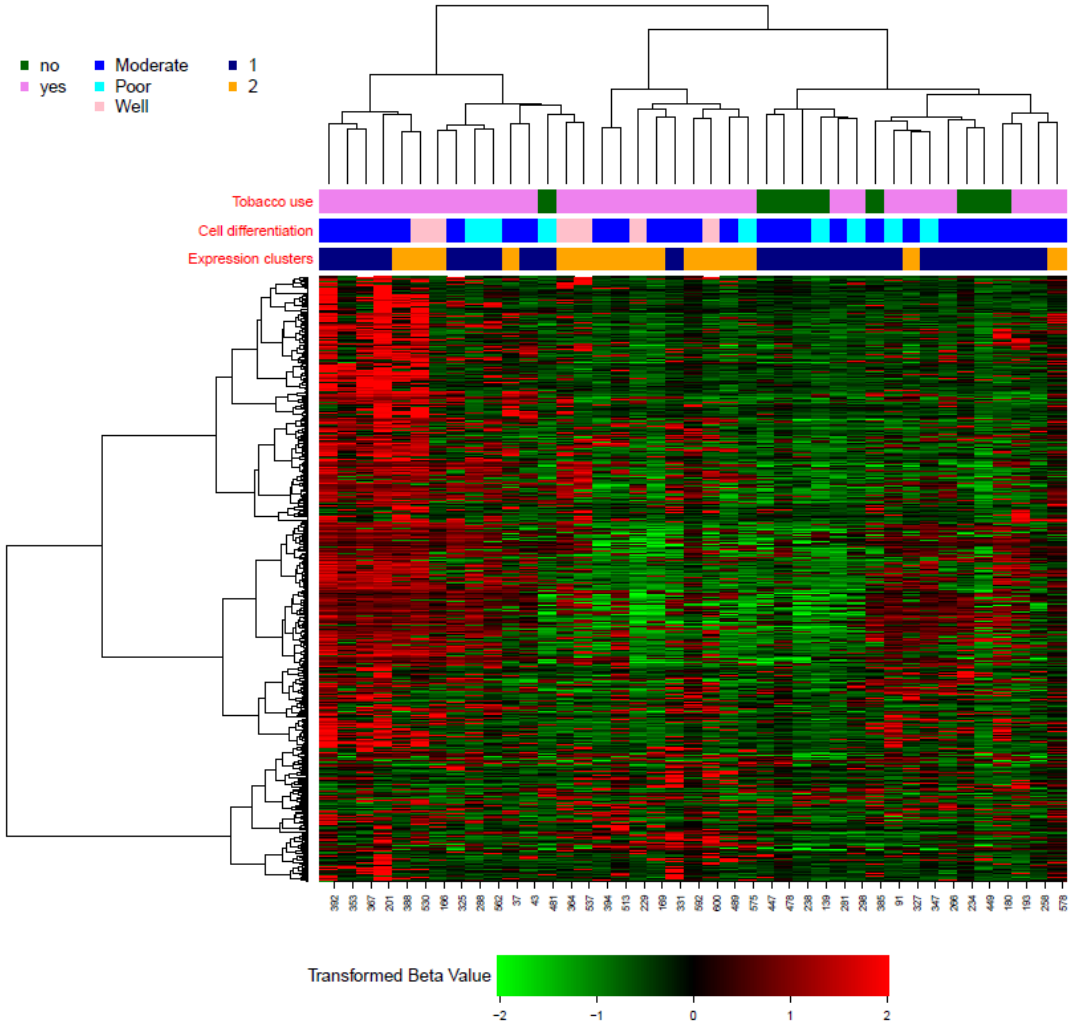

no
yes
Moderate
Poor
Well
1
2
Tobacco use
Cell differentiation
Expression clusters
Transformed Beta Value
−2
−1
0
1
2

Fig. SM10.

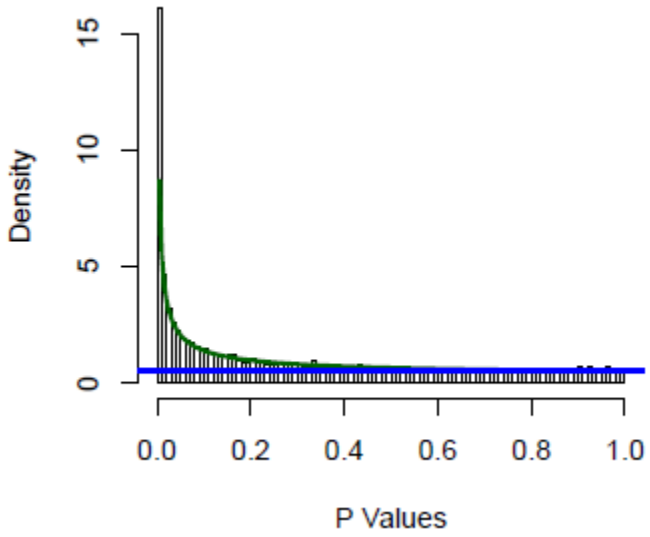


Fig. SM11.

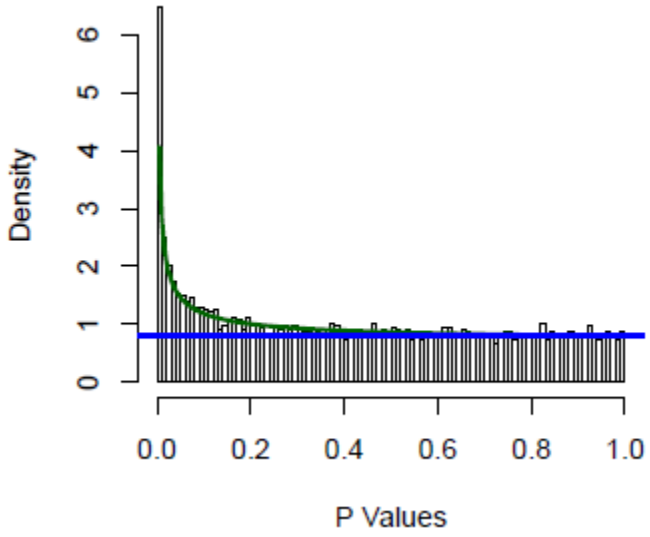


Fig. SM12.

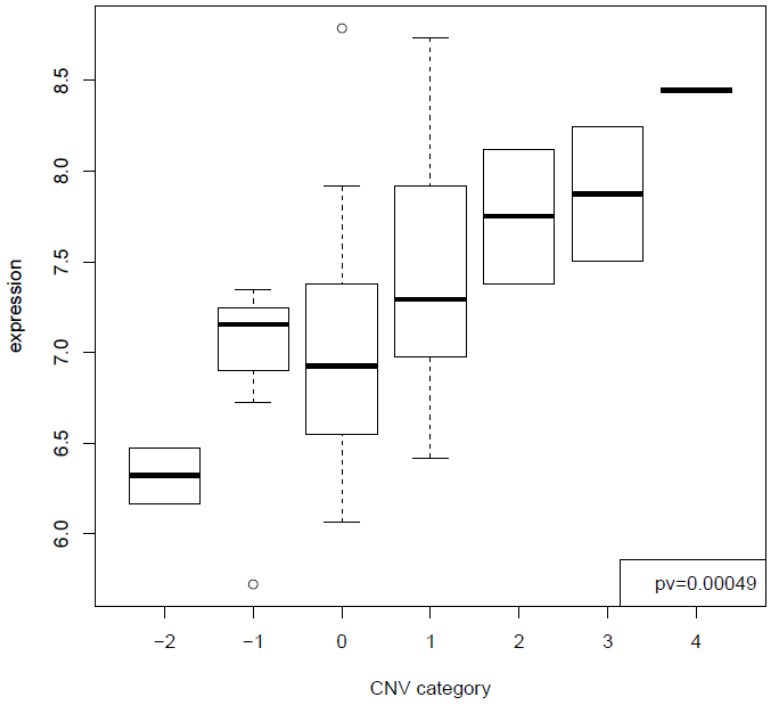


Fig. SM13.

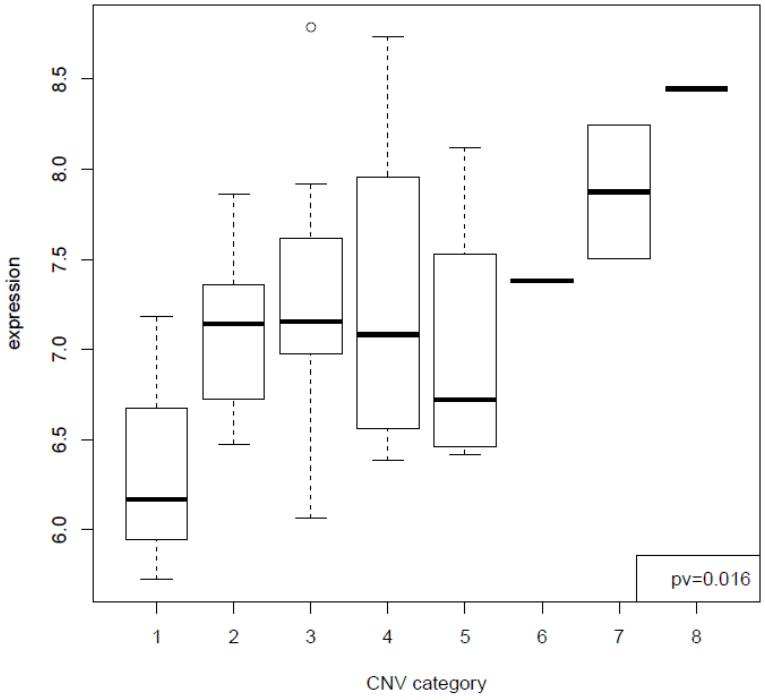
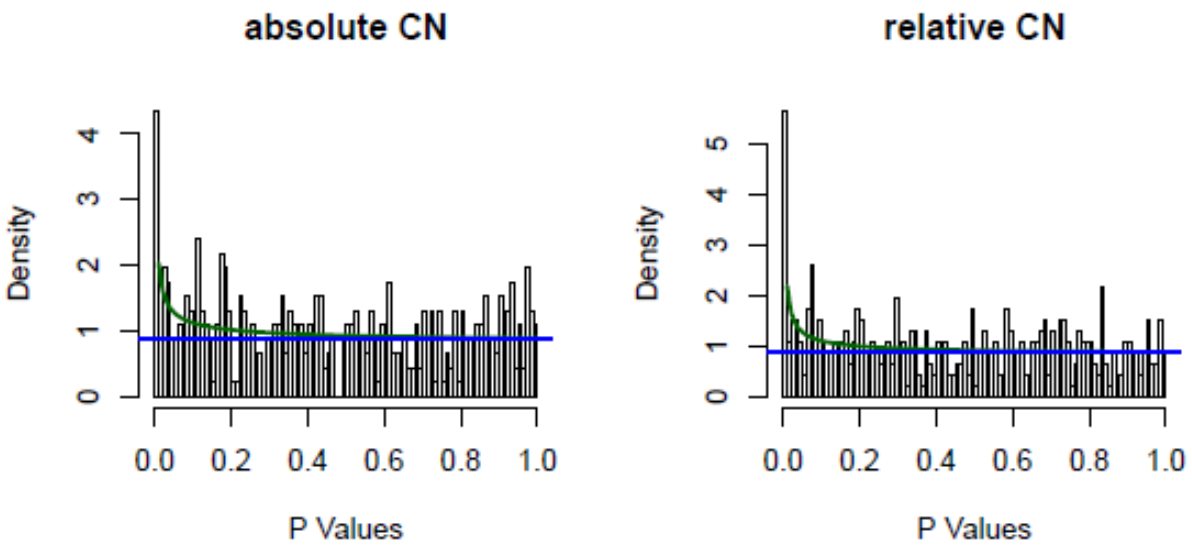


Fig. SM14.

	cn.loss	cn.unchanged	cn.gain
exp.down	6655	6119	2001
exp.normal	88227	371627	96063
exp.up	3088	12556	13418

Fig. SM15.


absolute CN
Density
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
P Values
relative CN
Density
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
P Values